

# Information flow between subspaces of complex dynamical systems

Andrew J. Majda<sup>†</sup> and John Harlim

Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012

Contributed by Andrew J. Majda, April 16, 2007 (sent for review March 21, 2007)

**The quantification of information flow between subspaces in ensemble predictions for complex dynamical systems is an important practical topic, for example, in weather prediction and climate change projections. Although information transfer between dynamical system components is an established concept for nonlinear multivariate time series, the specific nature of the nonlinear dynamics generating the observed flow of information is ignored in such statistical analysis. Here, a general mathematical theory for information flow between subspaces in ensemble predictions of a dynamical system is developed, which accounts for the specific underlying dynamics. The results below also include potentially useful approximation strategies for practical implementation in dynamical systems with many degrees of freedom. Specific elementary examples are developed here with both stable and unstable dynamics to both illustrate facets of the theory and to test Monte Carlo solution strategies.**

ensemble predictions | information transfer

It has been known for several decades that the practical prediction of complex chaotic dynamical systems such as those occurring in weather and climate requires a statistical framework involving ensembles of solutions (1–3). Sophisticated strategies for statistical ensemble generation and prediction in complex dynamical systems have been developed and applied to weather and climate prediction (4–10).

Recently, concepts from information theory (11, 12) have been introduced as natural tools to quantify the uncertainty in ensemble predictions for complex dynamical systems (13–21). For extended range forecasting or climate change projections, it is important to quantify the flow of information between subspaces of such complex dynamical systems in an ensemble prediction. For example, one might be interested in the impact of the subspace spanned by the atmospheric variables on the subspace spanned by the ocean variables in a climate change prediction or the impact of the subspace spanned by the high frequency synoptic scale eddies on the subspace of low frequency teleconnection patterns in an extended range ensemble forecast for the atmosphere (20). It is anticipated that such quantification of information flow between subspaces of complex dynamical systems will also prove to be useful in other diverse scientific disciplines such as neurology, nanotechnology, material science, and environmental science whenever ensemble predictions for the future state of a system are needed.

Information transfer between dynamical system components is an important established concept for coherence analysis of nonlinear multivariate time series (22–24); however, the specific nature of the nonlinear dynamics generating the observed flow of information is ignored in such statistical analysis. Recently, Liang and Kleeman (25) have put forward a new way to look at information flow between the two components of two-dimensional systems which addresses the change in uncertainty of one component due to the specific dynamic interaction with the second component. Here, their results are systematically generalized to a theory for the information flow between subspaces of complex nonlinear dynamical systems with many

degrees of freedom. In addition, several potentially practical estimation strategies for information flow for large dimensional complex nonlinear dynamical systems are developed here (16, 19, 21) as well as simple examples, illustrating various new phenomena. A completely different generalization of the work in ref. 25 is reported by those authors elsewhere (26, 27).

## Formulation and Basic Theory

Consider the state variable  $z$  for a nonlinear dynamical system so that  $z = (x, y)$  where  $x \in \mathbb{R}^N$ ,  $y \in \mathbb{R}^M$  are variables spanning two different subspaces. Here we assume that the dynamical system for  $z$  can be written as the equations

$$\begin{aligned}\frac{dx}{dt} &= F_1(x, y) \\ \frac{dy}{dt} &= F_2(x, y) + \sigma W,\end{aligned}\quad [1]$$

where  $\sigma\sigma^T = Q \geq 0$  is a nonnegative symmetric  $M \times M$  covariance matrix and  $W$  is  $M$ -dimensional white noise. Below, we consider the information flow in an ensemble prediction from the  $y$  variables to the  $x$  variables with noise to represent the role of additional unresolved processes on the dynamical system (12). The results below also apply to the important special case without noise, i.e.,  $\sigma \equiv 0$ .

An ensemble prediction for Eq. 1 defined through the probability density  $p(x, y, t)$  satisfies the Fokker–Planck equation

$$\begin{aligned}\frac{\partial p}{\partial t} + \nabla_x \cdot (F_1 p) + \nabla_y \cdot (F_2 p) &= \frac{1}{2} \nabla_y \cdot (Q \nabla_y p) \\ p(x, y, t)|_{t=0} &= p_0(x, y),\end{aligned}\quad [2]$$

where “ $\nabla_x \cdot()$ ” and “ $\nabla_y \cdot()$ ” denote divergences in  $x$  and  $y$  subspaces, respectively. Given a probability density,  $p$ , the absolute uncertainty in  $p$  is given by the entropy (11, 12)

$$H(p) = - \int p \ln p, \quad [3]$$

whereas the expected value of a random variable  $g$  with respect to  $p$  is denoted by  $\langle g \rangle_p = \int g p$ . For a probability density  $p(x, y)$ , the marginal probability distribution  $p_1(x)$  and the conditional distribution  $p(y|x)$  are defined as

$$p_1(x) = \int p(x, y) dy, \quad p(x, y) = p_1(x) p(y|x) \quad [4]$$

with  $\int p(y|x) dy = 1$ .

Author contributions: A.J.M. designed research; A.J.M. and J.H. performed research; and A.J.M. and J.H. wrote the paper.

The authors declare no conflict of interest.

<sup>†</sup>To whom correspondence should be addressed. E-mail: jonjon@cims.nyu.edu.

© 2007 by The National Academy of Sciences of the USA

First, we state an intuitive definition for information flow between subspaces and then we translate this definition into a precise mathematical statement and supply an exact formula for the evolution of this quantity

**Definition 1.** For the statistical solution  $p(x, y, t)$  satisfying Eq. 2, the *information flow* at time  $t$  from the subspace spanned by  $y$  to the subspace spanned by  $x$ ,  $T_{y \rightarrow x}(p)$ , is the time rate of change of the difference in the absolute uncertainty in the  $x$  variables due to the dynamic information in the  $y$  variable.

The absolute uncertainty in the  $x$  variables alone at time  $t$  is given by  $H(p_1)$  where  $p_1(x, t)$  is the marginal distribution of  $p(x, y, t)$ . To isolate the dynamic information at time  $t$  in the  $y$  variables, consider the dynamical system related to Eq. 1 where only the  $x$  variables evolve nontrivially, i.e.,  $dx/dt = F_1$ ,  $dy/dt = 0$ ; the corresponding Fokker–Planck equation for a probability density  $p^*(x, y, t)$  for  $t > \tau$  in this related system with initial density  $p^*(x, y, \tau) = p(x, y, \tau)$  is given by

$$\begin{aligned} \frac{\partial p^*}{\partial t} + \nabla_x \cdot (F_1 p^*) &= 0 \\ p^*(x, y, t)|_{t=\tau} &= p(x, y, \tau). \end{aligned} \quad [5]$$

It is easy to see that  $p^*(x, y, t) = p^*(x|y, t) p_2(y, \tau)$  where  $p_2(y, \tau) = \int p(x, y, \tau) dx$  is the  $y$  marginal and also  $p^*(x|y, t)$  satisfies the same Eq. 5. Thus,  $p^*(x, y, t)$  for  $t > \tau$  expresses the ensemble evolution which could occur through the  $x$  variables alone with the probability distribution for the  $y$  variables frozen at the density  $p_2(y, \tau)$ . The quantity  $H(p^*)$  for  $t > \tau$  measures the absolute uncertainty in the  $x$  variables due to the  $x$  dynamics alone. With these preliminaries, following ref. 25, the information flow at time  $t$  from the subspaces  $y$  to the subspace  $x$  is given by

$$T_{y \rightarrow x}(p) = \frac{d}{dt} H(p_1) - \frac{d}{dt} H(p^*)|_{t=\tau} \quad [6]$$

The following is a general theoretical formula which takes into account the specific nonlinear dynamics in Eq. 1 in the information flow

**Theorem 1.**

$$T_{y \rightarrow x}(p) = - \int \int p(y|x, t) \nabla_x \cdot (F_1(x, y) p_1(x, t)) dx dy \quad [7]$$

Note that the information flow is a flux into the  $x$  variables averaged with respect to the conditional distribution  $p(y|x, t)$ . The proof of this Theorem is a direct calculation. First, we readily compute

$$\frac{d}{dt} H(p^*)|_{t=\tau} = \int \int (p \nabla_x \cdot F_1(x, y)) dx dy \quad [8]$$

Second, we calculate that

$$\begin{aligned} \frac{d}{dt} H(p_1) &= - \int \int \frac{dp}{dt} (\ln p_1 + 1) dx dy \\ &= \int \int \nabla_x \cdot (F_1 p) (\ln p_1 + 1) dx dy \end{aligned}$$

$$\begin{aligned} &+ \int \left( \int \nabla_y \cdot \left( F_2 p - \frac{1}{2} Q \nabla p \right) dy \right) (\ln p_1 + 1) dx. \\ &= - \int \int p(y|x, t) (F_1 \cdot \nabla p_1(x, t)) dx dy. \end{aligned} \quad [9]$$

The difference between Eqs. 9 and 8, decomposed through Eq. 4, yields the formula in Eq. 7. Note that, in the final equality Eq. 9, the integration by parts and Eq. 4 have been used.

Next, assume that the nonlinearity,  $F_1(x, y)$  has the decomposition

$$F_1(x, y) = F_{11}(x) + F_{12}(x, y). \quad [10]$$

From Eq. 4, it follows that

$$\begin{aligned} &\int \int p(y|x, t) \nabla_x \cdot (F_{11}(x) p_1(x, t)) dx dy \\ &= \int \nabla_x \cdot (F_{11}(x) p_1(x, t)) dx = 0. \end{aligned}$$

Thus, the following result is immediate that confirms our intuition about information flow (25):

**Proposition 2.** If  $F_1$  has the decomposition in Eq. 10, then  $F_{11}(x)$  does not contribute to the information flow,  $T_{y \rightarrow x}(p)$ .

An equilibrium distribution  $p_{eq}(x, y)$  is a probability density which is a time-independent solution of the Fokker–Planck equation in Eq. 2. We have the following general fact:

**Proposition 3.** If  $p_{eq}(x, y)$  factors into the product distribution,  $p_{eq}(x, y) = p_{eq,1}(x) p_{eq,2}(y)$ , then there is no information flow at equilibrium, i.e.,  $T_{y \rightarrow x}(p) = 0$ .

Under the assumption in Proposition 3, we have

$$\begin{aligned} T_{y \rightarrow x}(p_{eq}) &= - \int \int p_{eq,2}(y) \nabla_x \cdot (p_{eq,1}(x) F_1(x, y)) dx dy \\ &= \int \int (F_1 \cdot \nabla_x p_{eq,2}(y)) p_{eq,1}(x) dx dy = 0. \end{aligned}$$

So, Proposition 3 follows immediately and the intuition is confirmed that if the equilibrium distribution factors into a product in the  $x$  and  $y$  variables, then there is no information flow.

The next set of identities for information flow are useful for developing practical computational strategies. First, assume that with a smooth function  $\phi(x, t)$ , the marginal distribution  $p_1(x, t)$  has the form

$$p_1(x, t) = \exp(-\phi(x, t)) \quad [11]$$

and that the nonlinear function  $F_1$  has the decomposition in Eq. 10. Under these assumptions, we have the following identity for information flow

**Proposition 4.**

$$T_{y \rightarrow x}(p) = -\langle \nabla_x \cdot F_{12} \rangle_p + \langle F_{12} \cdot \nabla_x \phi \rangle_p \quad [12]$$

The result in Eq. 12 follows immediately from Proposition 2 and the following calculus identity using Eqs. 11 and 12

$$p(y|x, t) \nabla_x \cdot (F_{12} p_1) = p \nabla_x \cdot F_{12} - p F_{12} \cdot \nabla_x \phi.$$

There is a formula similar to Eq. 12 under the assumption that the conditional probability distribution satisfies a parallel assumption as in Eq. 11, i.e., there is a smooth function  $\psi(x, y, t)$  so that

$$p(y|x, t) = \exp(-\psi(x, y, t)). \quad [13]$$

Under the assumption in Eq. 13, the following formula is valid for the information flow

**Proposition 5.**

$$T_{y \rightarrow x}(p) = -\langle F_{12} \cdot \nabla_x \psi \rangle_p. \quad [14]$$

First, one integrates by parts in Eq. 7 and then uses the form in Eq. 13 with a similar calculus identity as developed below Eq. 11 to establish Eq. 14. To illustrate the potential use of Proposition 4, assume that the marginal distribution is Gaussian, i.e., assume

$$p_1(x, t) = c_o(t) \exp\left(-\frac{1}{2} (x - \langle x \rangle_p)^T \Gamma^{-1}(t) (x - \langle x \rangle_p)\right),$$

where  $\Gamma(t)$  is the covariance matrix, then from Proposition 4

$$T_{y \rightarrow x}(p) = -\langle \nabla_x \cdot F_{12} \rangle_p + \langle F_{12} \cdot \Gamma^{-1}(t) (x - \langle x \rangle_p) \rangle_p. \quad [15]$$

**Prototype Geophysical Models and Information Flow Relative to the Climatology**

One of the important potential applications of the above theory is to information flow in ensemble predictions for weather prediction and climate change projection. As discussed, for example, in refs. 11 and 12, the dynamical core of atmosphere/ocean general circulation models as well as a large family of intermediate complexity models can be written as quadratically nonlinear dynamical systems,

$$\frac{dz}{dt} = Lz + B(z, z) - Dz + F, \quad [16]$$

where  $L$  and  $B$  satisfy

$$L^T = -L, \quad z \cdot B(z, z) = 0 \quad [17]$$

and thus conserve energy, and  $D$  and  $F$  represent damping and forcing. Introduce the decomposition,  $z = (x, y)$  and assume special random forcing and dissipation in the  $y$  variable alone as in Eq. 1 to define the prototype system

$$\frac{dx}{dt} = (L_{11}x + B_{11}^1(x, x)) + (L_{12}y + B_{12}^1(x, y) + B_{22}^1(y, y))$$

$$\frac{dy}{dt} = (L_{21}x + B_{11}^2(x, x) + B_{21}^2(x, y)) + (L_{22}y + B_{22}^2(y, y))$$

$$- dy + \sigma W. \quad [18]$$

Here,  $d \geq 0$  is dissipation in the  $y$  variables. Besides Eq. 17, it is natural to assume (see refs. 11 and 12) that the nonlinear terms satisfy

$$\begin{aligned} x \cdot B_{11}^1(x, x) &= 0, & y \cdot B_{22}^2(y, y) &= 0 \\ \nabla_x \cdot B_{11}^1(x, x) &= 0, & \nabla_y \cdot B_{22}^2(y, y) &= 0 \\ \nabla_x \cdot B_{12}^1(x, y) &= 0, & \nabla_y \cdot B_{22}^2(y, y) &= 0. \end{aligned} \quad [19]$$

There are many examples of geophysical systems satisfying Eqs. 17 and 19, including all of the multilayer models in section 1.6 of refs. 11 and 12 as well as important geophysical test models such as the Lorenz-96 model (11, 12, 19, 21), the truncated Burgers-Hopf and the Kruskal-Zabusky models (11, 12, 17).

First, ignore the dissipation and forcing in Eq. 18; then it follows from Eqs. 17 and 19 that the system in Eq. 18 has the family of Gaussian invariant measures

$$p_{eq} = C_\sigma \exp\left(-\frac{1}{2\sigma^2} (x \cdot x + y \cdot y)\right), \quad [20]$$

which clearly satisfy Proposition 3 above. Furthermore, it was shown in section 6 of ref. 28 that Eq. 20 remains as the invariant measure with appropriate dissipation and white noise forcing.

The relative entropy of a probability density  $p$  with respect to the prior density  $\pi$  is given by the nonnegative functional

$$P(p, \pi) = \int p \ln\left(\frac{p}{\pi}\right). \quad [21]$$

It measures the average lack of information in  $\pi$  relative to  $p$  (see refs. 11 and 12); Kleeman (15) has introduced the relative entropy of an ensemble prediction  $p$ , relative to the climatology,  $\pi$ , as a measure of utility in an ensemble prediction and this notion has been developed extensively (11, 12, 16–18, 20, 21). In the present context, the relative information in the marginal compared with the marginal of the climatology in Eq. 20 is the object of interest,

$$\begin{aligned} P(p_1(x, t), p_{eq}, 1(x)) &= -H(p_1(x, t)) - \ln C_\sigma \\ &+ \frac{1}{2\sigma^2} \int (x \cdot x) p_1(x, t) dx. \end{aligned} \quad [22]$$

In the present models, Eqs. 8 and 19 guarantee  $dH(p^*)/dt = 0$ , whereas

$$E_x(t) = \frac{1}{2} \int (x \cdot x) p_1(x, t) dx \quad [23]$$

is the amount of the energy which resides in the subspace spanned by  $x$  at time  $t$  during the ensemble prediction. With Eqs. 2, 17, 18, 19, and 23, one readily calculates the identity

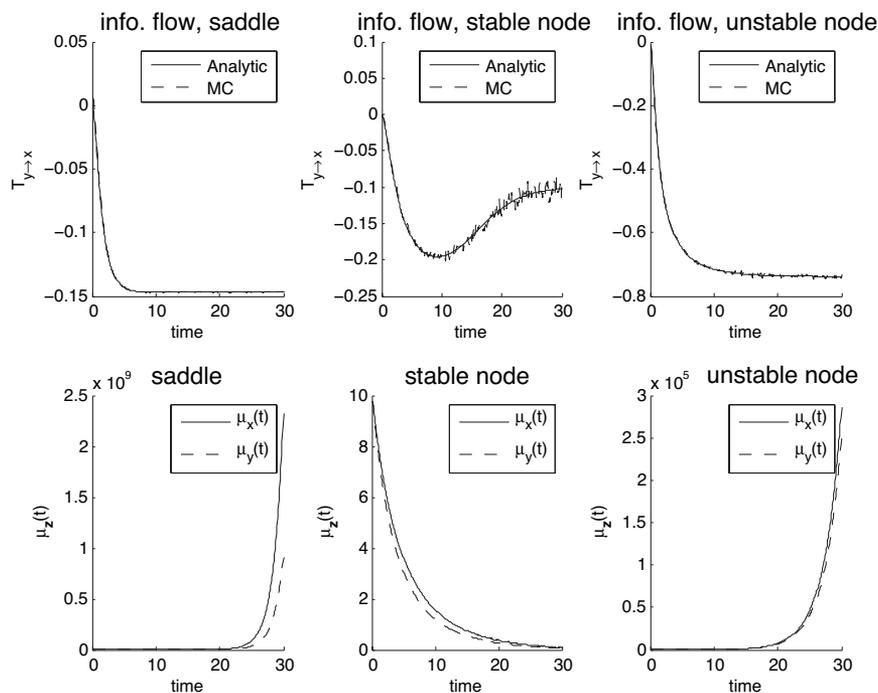
$$\frac{d}{dt} E_x(t) = \langle x \cdot (L_{12}y + B_{12}^1(x, y) + B_{22}^1(y, y)) \rangle_p \equiv E_{y \rightarrow x}(p) \quad [24]$$

for the rate of energy transfer from the  $y$  variables. Using all of the facts below Eq. 22, we have the following identity for information flow relative to the climatology

**Proposition 6.**

$$\frac{d}{dt} P(p_1, p_{eq,1}) = -T_{y \rightarrow x}(p) + \sigma^{-2} E_{y \rightarrow x}(p). \quad [25]$$

Thus, under the hypothesis of Eq. 18, we have directly related information flow relative to the climatology in the variables  $x$  in an ensemble prediction to  $T_{y \rightarrow x}(p)$  and the rate of energy transfer  $E_{y \rightarrow x}(p)$  for the ensemble prediction. Clearly, the optimal value of  $\sigma$  in Eq. 20 for comparison with the ensemble simulation is the value with the same mean energy. Proposition 6 relates the information flow,  $T_{y \rightarrow x}(p)$ , to the time rate of change of utility in an ensemble forecast and serves as a starting point for further generalizations.



**Fig. 1.** Nonoscillatory model. (Left) Solution with fixed point of saddle type. (Center) Stable node. (Right) Unstable node. (Upper) The information flow  $T_{y \rightarrow x}$  as a function of time. (Lower) Ensemble mean states  $\mu_x(t)$  in solid and  $\mu_y(t)$  in dashes, both, as functions of time.

### An Instructive Model Demonstrating Information Flow

Here, we briefly illustrate several facets of information flow in an ensemble prediction for the  $2 \times 2$  system of linear stochastic equations

$$\begin{aligned} \frac{dx}{dt} &= \alpha x - \beta y \\ \frac{dy}{dt} &= \beta x - \gamma y + \sigma \dot{W}. \end{aligned} \quad [26]$$

With a Gaussian initial data, the above linear equation has a Gaussian solution for all times that is readily computed exactly by explicit formulas (29). Furthermore, *Proposition 4* and Eq. 15 apply for all times so that the information flow in these examples is given by

$$T_{y \rightarrow x}(p) = -\beta \frac{\Gamma_{12}(t)}{\Gamma_{11}(t)}, \quad [27]$$

with  $\Gamma_{12}(t) = \langle (x - \langle x \rangle_p)(y - \langle y \rangle_p) \rangle_p$  the cross-covariance of  $x$  and  $y$  and  $\Gamma_{11}(t) = \langle (x - \langle x \rangle_p)^2 \rangle_p$  the variance of  $x$ .

Below, the value of the damping coefficient  $\gamma$  is set to one, whereas the noise level is fixed by  $\sigma = 0.2$ ; the parameters  $\alpha$  and  $\beta$  are varied systematically so that the mean of the solution of Eq. 26 exhibits complete variety of stable and unstable behavior ranging from a saddle to stable and unstable nodes to stable and unstable spirals to a Hamiltonian center. In each of these six cases, an ensemble prediction is generated through a 1,000 member ensemble with initial ensemble mean (10, 10) and uncorrelated unity covariance matrix. In the ensemble prediction, the information transfer in Eq. 27 is approximated by the Monte Carlo formula

$$T_{y \rightarrow x} \approx -\beta \frac{\sum_{i=1}^K (x_i(t) - \mu_x)(y_i(t) - \mu_y)}{\sum_{i=1}^K (x_i(t) - \mu_x)^2},$$

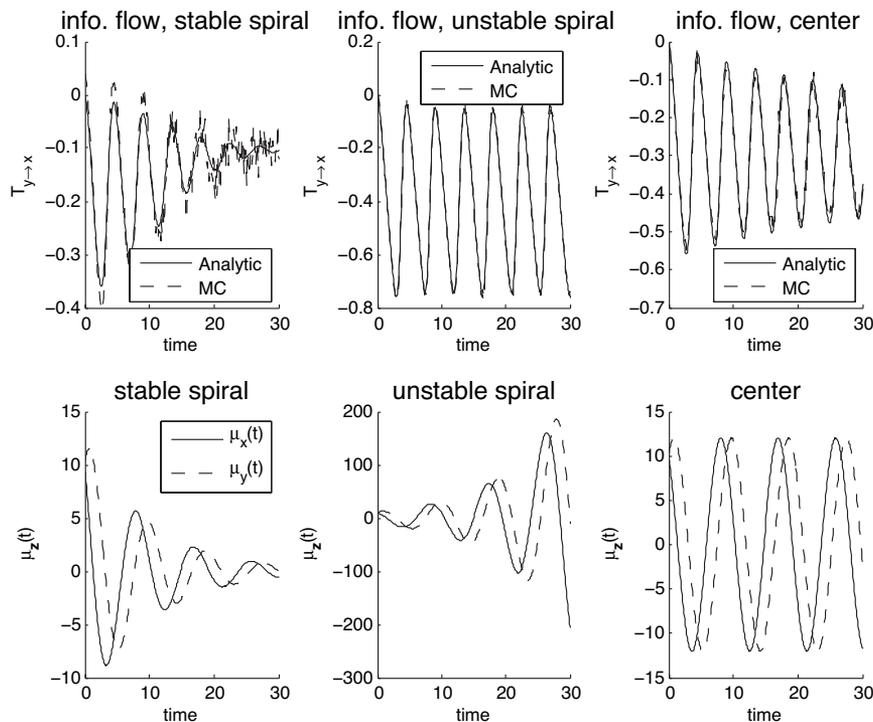
with  $\mu_x$  and  $\mu_y$  denote the Monte Carlo ensemble average over variables  $x$  and  $y$  for ensemble size of  $K = 1,000$ , respectively.

The comparison of the exact analytic formulas and the Monte-Carlo ensemble simulation of information transfer for the six cases are depicted in Figs. 1 and 2. Fig. 1 *Upper* shows the information transfers as functions of time from left to right for a saddle, stable node, and unstable node, whereas Fig. 2 *Upper* is for a stable spiral, unstable spiral, and Hamiltonian center. Figs. 1 *Lower* and 2 *Lower* show the ensemble average of the dynamical variables  $x$  and  $y$  as functions of time. The overall trend of the information transfer in all of these diverse cases is captured very well by the Monte Carlo simulations; this is even true for the cases with a positive Lyapunov exponent and no invariant measure involving the saddle, unstable node, and unstable spiral even though the mean has grown to the respective values  $2 \times 10^9$ ,  $2 \times 10^5$ , and  $2 \times 10^2$  on the time interval depicted.

There is no oscillation in the mean in the cases in Fig. 1, whereas there is decaying, amplifying, and neutral oscillation of the mean state, respectively, in Fig. 2, and such oscillatory behavior is clearly reflected in the information transfer  $T_{y \rightarrow x}(p)$ . More interestingly, the frequency of the oscillation in the information flow roughly doubles the frequency of the mean state (or signal). This is easy to understand because the random fields  $x(t)$  and  $y(t)$  from the explicit solution both contain pieces with the same temporal frequency  $\omega$ ; thus, the quadratic covariance from the information transfer in Eq. 27 has a contribution with frequency  $2\omega$ , i.e., one half the period. Finally, one should note that in these examples, especially for those with equilibrium density  $p_{eq}$  (e.g., the stable node, the stable spiral, and the Hamiltonian center), the information flow never converges to zero as suggested in *Proposition 3* because the cross-covariance between  $x$  and  $y$  is nonzero at equilibrium.

### Systematic Approximations for Information Flow

A theory for information flow in ensemble predictions has been outlined above but the practical utility of such ideas rests on the



**Fig. 2.** Oscillatory model. (Left) Solution with fixed point of stable spiral. (Center) Unstable spiral. (Right) Center. (Upper) The information flow  $T_{y \rightarrow x}$  as a function of time. (Lower) Ensemble mean states  $\mu_x(t)$  in solid and  $\mu_y(t)$  in dashes, both, as functions of time.

ability to develop practical approximate algorithms for information flow. This is the topic here. Two general methods are proposed below.

The first method builds on the general formalism in Eq. 11 and Proposition 4 for information flow. First, assume that the general marginal distribution,  $p_1(x, t)$ , is approximated by

$$p_1(x, t) \cong p_{1,J}(x, t) = \exp(-\phi_J(x, t)),$$

with

$$\phi_J(x, t) = \sum_{j=1}^J \lambda_j x^j$$

where  $\phi_J(x, t)$  is chosen to be the least biased probability density which matches the mean and higher moments of  $p_1(x, t)$  up to order  $J$ . This is the strategy used in refs. 16, 17, and 19–21 for computing relative information in ensemble predictions, and there are efficient algorithms to compute this rapidly provided  $x$  is one or two-dimensional (19–21, 30); for the prototype geophysical models in Eq. 16, the formula in Eq. 12 of Proposition 4 can be evaluated explicitly using only ensemble moments. The simplest version of this algorithm occurs when  $J = 2$  so that  $\phi_J(x, t)$  defines a Gaussian approximation; the needed approximate formula is already listed in general in Eq. 15 and of course, the advantage of the Gaussian approximation is that it applies when  $x$  has an arbitrary number of variables. For the truncated Burgers–Hopf model, it has been established (see ref. 17) that, surprisingly, typical ensemble predictions are nearly Gaussian so for this system, such an approximation would be excellent. Alternatively, the same approximation strategy can be applied to the conditional distribution  $p(y|x, t)$  by using suitable similar approximations to  $\psi$  in Eq. 13 and Proposition 4.

The second strategy for computing information flow in an ensemble prediction involves asymptotic methods that exploit the disparity in time scales between the  $x$  and  $y$  processes for special systems such as

$$\begin{aligned} \frac{dx}{dt} &= F_1(x, y) \\ \frac{dy}{dt} &= \varepsilon^{-1} F_2(x, y) + \varepsilon^{-1/2} \sigma \dot{W} \end{aligned} \quad [28]$$

with  $\varepsilon \ll 1$ . In this case, one assumes ergodicity and a unique smoothly varying invariant measure,  $p_{\text{eq}}(y|x)$ , for the reduced system

$$\frac{dy}{d\tau} = F_2(x, y) + \sigma \dot{W}$$

with  $\tau = t/\varepsilon$  and  $x$  regarded as a parameter.

With these and other suitable assumptions, it is not too difficult to establish rigorously (see ref. 31 for examples) that as  $\varepsilon \rightarrow 0$  for any positive time,

$$p_\varepsilon(x, y, t) = p_1(x, t) p_{\text{eq}}(y|x) + o(1), \quad [29]$$

so in this setting for  $\varepsilon \ll 1$ , the information transfer is given approximately by

$$T_{y \rightarrow x}(p) \cong - \iint p_{\text{eq}}(y|x) \nabla_x \cdot (F_1 p_1) dx dy. \quad [30]$$

The battery of approximate procedures discussed in the previous paragraph can be applied to  $p_{\text{eq}}(y|x)$  in the present context. A more attractive computational option would involve

using “on the fly” computational procedures for Eq. 28 (see ref. 32).

### Concluding Discussion

A systematic mathematical theory for information transfer between subspaces in ensemble predictions of nonlinear dynamical systems has been developed here including potential practical computational strategies and a simple illustrative test example. Further

comprehensive numerical tests are needed to demonstrate the practical utility of calculating information transfer in realistic complex dynamical systems.

The research of A.J.M. is partially supported by the National Science Foundation, the Office of Naval Research, and the Defense Advanced Research Projects Agency. J.H. is supported as a postdoctoral fellow through the last two agencies.

1. Leith CE (1974) *Mon Weather Rev* 102:409–418.
2. Lorenz EN (1965) *Tellus* 17:321–333.
3. Epstein ES (1969) *Tellus* 21:739–759.
4. Buizza R, Palmer T (1998) *Mon Weather Rev* 126:2503–2518.
5. Ehrendorfer M, Tribbia J (1997) *J Atmos Sci* 54:286–313.
6. Kalnay E (2003) *Atmospheric Modeling, Data Assimilation, and Predictability* (Cambridge Univ Press, Cambridge, UK), p 363.
7. Palmer T (2000) *Rep Prog Phys* 63:71–116.
8. Palmer T, Molteni F, Mureau R, Buizza R, Chapelet P, Tribbia J (1993) *Proceedings of the ECMWF Seminar on Validation Models Over Europe* (ECMWF, Shinfield Park, Reading, UK), Vol 1, pp 21–66.
9. Toth Z, Kalnay E (1993) *Bull Am Meteor Soc* 74:2317–2330.
10. Toth Z, Kalnay E (1997) *Mon Weather Rev* 125:3297–3319.
11. Majda AJ, Wang X (2006) *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flow* (Cambridge Univ Press, Cambridge, UK).
12. Majda AJ, Abramov R, Grote M (2005) *Information Theory and Stochastic for Multiscale Nonlinear Systems* (Am Math Soc, Washington, DC), CRM monograph series 25.
13. Carnevale G, Holloway G (1982) *J Fluid Mech* 116:115–121.
14. Schneider T, Griffies SM (1999) *J Climate* 12:3133–3155.
15. Kleeman R (2002) *J Atmos Sci* 59:2057–2072.
16. Majda AJ, Kleeman R, Cai D (2002) *Meth Appl Anal* 9:425–444.
17. Kleeman R, Majda AJ, Timofeyev I (2002) *Proc Natl Acad Sci* 99:15291–15296.
18. Kleeman R, Majda AJ (2006) *J Atmos Sci* 62:2864–2879.
19. Haven K, Majda AJ, Abramov R (2004) *J Comp Phys* 206:334–362.
20. Abramov R, Majda AJ, Kleeman R (2005) *J Atmos Sci* 62:65–87.
21. Abramov R, Majda AJ (2004) *SIAM J Sci Stat Comp* 26:411–447.
22. Vastano JA, Swinney HL (1988) *Phys Rev Lett* 60:1773–1776.
23. Schreiber T (2000) *Phys Rev Lett* 85:461.
24. Kaiser A, Schreiber T (2002) *Physica D* 166:43–62.
25. Liang XS, Kleeman R (2005) *Phys Rev Lett* 95:244101–244104.
26. Liang XS, Kleeman R (2007) *Physica D*, in press.
27. Liang XS, Kleeman R (2006) *Physica D* 227:173–182.
28. Majda AJ, Timofeyev I, Vanden-Eijnden E (2001) *Comm Pure Appl Math* 54:891–974.
29. Gardiner CW (1985) *Handbook of Stochastic Methods* (Springer, Berlin).
30. Abramov R (2006) *J Comp Phys* 211:198–209.
31. Imkeller P, von Storch J-S (2001) *Proceeding of the Workshop on Stochastic Climate Models* (Birkhäuser, Boston).
32. Vanden-Eijnden E (2003) *Comm Math Sci* 1:385–391.